

## Minimal Data Storage Requirements NIN

August 2018 – draft by NIN Data Committee,

Updated November 28th, 2018

Updated December 10<sup>th</sup>, 2021 by NIN Data Committee

**From April 2<sup>nd</sup>, 2019, each NIN researcher needs to comply with the following storage protocol. Compliance will be checked randomly by the NIN Data Officer and the results of these checks will be submitted to the board.**

**Electronic research data** contains raw data and associated meta-data, but also includes experiment scripts, experimental configuration parameters, experimental log files, data analysis scripts, the relevant ethics protocol and approvals, submitted manuscripts.

**Data storage.** Before publication, all electronic research data collected from April 2<sup>nd</sup> 2019 onwards should at least be stored at a NIN-consented safe-storage location. NIN-consented safe-storage location are backed-up storage locations which have the explicit consent of the NIN Data Officer, and are readable by at least the responsible group leader and NIN Data officer. Two examples of the NIN-consented safe-storage locations are folders on the NIN's backed-up file server and folders on group Dropbox for Business accounts.

**Who stores?** All NIN staff members (group leaders, post docs and PhD-students) who are lead scientists for a study are responsible for appropriate data storage. BA and Master students should be taught to follow the storage rules, but their daily supervisor is responsible that the data is properly stored, if these data become part of a publication.

**Publication.** When the results of a data set are published, the first author is responsible for uploading the data to a NIN-consented archiving location maximally 1 month after (online) publication. This obligation holds even if the first author is then no longer at the NIN. NIN-consented archiving locations are backed-up storage locations which can hold immutable data and have the explicit consent of the NIN Data Officer. This location is readable by at least the responsible group leader and NIN Data officer. An example of such storage is the read-only Publications folder on the NIN file server maintained by the NIN Data Officer. Archiving the data at this location does not automatically make the data publicly accessible.

**Storage duration.** All (raw) data used in a publication should be stored for at least 10 years after the publication date.

**Anonymity.** Data should be stored in such a way that it is not possible to track down personal details from the data. This means that data should be pseudonimized. In practice, this means that:

- a. All informed consent forms should be scanned and encrypted and the paper copy archived.
- b. Encryption is required at least for sensitive data (e.g. patients), video data of humans (unless permission to distribute publicly has been obtained), and questionnaires which include personal data.

**Data belonging to other parties.** If data is owned by other organizations, other universities, or persons, who would object to storage, then this should be documented in the project folder.

**Data structure.** On the next page, a **structure** is recommended for research project data folders (based on the UvA FMG structure). If a different structure is used, then a description of the data structure should be included with the data. In future, this recommended structure may become the required structure.

**Data sharing.** After publication, we recommend also uploading the data to an online repository, such as DANS, OSF (<https://osf.io>), Zenodo, or another on <https://www.nature.com/sdata/policies/repositories>.

## STRUCTURE AND CONTENTS OF A RESEARCH PROJECT FOLDER

The name of the folder should be unique for the research project. For any publication, it should be easy to identify in which research project folder the data is stored.

### MAINFOLDER

The main folder should contain a readme file with a short description of the project, including who is responsible and collaborating in the project. This folder should also contain a description of the data management plan.

### SUBFOLDERS

#### Ethics

- Ethics Protocol and Approval (PDF exported from EC site, METc or DEC, CCD or IvD)
- Information brochure, materials and debriefing brochure (as uploaded in an EC submission).

#### Methods\_and\_materials

- The **experiment scripts** (e.g., the Presentation, Eprime task, Matlab code) and stimuli
- PDF of paper-pencil questionnaires.
- If online survey software is used (e.g., Qualtrics), it is often possible to export the questions in PDF. Always include a description of the *exact version* and *platform* of the software used. For proprietary software (written expressly for an experiment, e.g., by TOP), a zip file with the software itself may be uploaded with installation instructions and a clear mention of the platform (e.g., Windows XP or higher) and machine (e.g., Esprima PC).

#### Data\_collection

- Appropriate configuration parameters (e.g., EEG configuration files, MRI protocols)
- Metadata with entries identified by date and experimenter; subjects identified by participant number, any subject-related materials or comments
- *Raw* data files (e.g., Presentation output files, rec/par, microscope images, etc.)

#### Data\_analysis

- All **scripts and syntax-files** used to transform and/or analyze the data (e.g., Excel files, SPSS.sav and syntax files, EEG and MRI analysis scripts).
- A file noting details about the data analysis of specific subjects relevant for replication
- A list of dropped subjects plus reason for exclusion
- A *code book*: description of all variable names and labels with sufficient detail to understand both the raw and processed data
- The transformed **data** that formed the basis for all analyses in the published paper.

#### Publications

- The final, submitted manuscript (PDF, refer to as final publication).

## Appendix December 2021

### Responsibilities

The researcher or data collector is the person that performs data management. This includes storing, documenting and organizing data according to the prescribed guidelines set out in this document. The researcher will be held responsible for making sure research data management is adequately performed.

The data officer/steward has the responsibility and right to check whether researchers realize their research data management according to these standards. The data officer should have the means and opportunity to question and investigate the data management practices of researchers and if necessary give advice or demand steps for improvement.

The Institute and group leaders are responsible for ensuring that researchers have enough time, resources and training to perform adequate research data management. Group leaders are required to stimulate and support appropriate data management of their group members.

### The data management plan

The data management plan is a document that changes as a project develops and should be updated to reflect current data storage requirements and data organization.

At its conception, *i.e. before data is actually collected*, this document should contain a description of :  
How and where data will be stored and how much storage space will be required.

How the data will be organized, and what file types will be saved.

If necessary, whether sensitive data will be collected and how this will be encrypted, and securely stored.

Where the data will be archived in the future for long term storage.

(see also : <https://www.nwo.nl/en/research-data-management> )

### The data collection folder

To make data sharing easier and enhance collaboration, data needs to be published in a recognizable and interoperable format. This requires standardization of data organization and metadata formats.

The following structure (naming schema) is recommended. Other schemes that are used in specific fields, can also be suitable (e.g. BIDS):

(Root folder) - **ProjectId**

(subfolder) - **DatasetId**

(subfolder) - **SubjectId**

(subfolder) - **Date\_sessionnr**

preprocessed/derived files

subjectid\_date\_###\_session.json

(subfolder) - **Raw**

raw data files

### Registration of data

At the NIN a system has been developed called "Follow Your Data" to register raw data files. This is done by placing a small specifically prescribed metadata file (json) in each separate data folder (see above), which is then automatically indexed and accessible from our database. For experimental data stored on the local NIN storage servers (VS01, VS03, VS- ) usage of Follow Your Data is recommended.

Researchers using Follow your Data, are also required to add documentation about their projects and datasets through the webapp associated with this system.

(Guidelines for organizing the data collection folder and using Follow Your Data are on :  
<https://hereninstituut.sharepoint.com/sites/RDM/SitePages/FAIR-data.aspx> )